Cell
Leading Edge
Minireview

No-Nonsense Functions for Long Noncoding RNAs

Takashi Nagano[1] and Peter Fraser[1,*]
[1]Nuclear Dynamics and Function Laboratory, The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK
*Correspondence: peter.fraser@bbsrc.ac.uk
DOI 10.1016/j.cell.2011.03.014

The mysterious secrets of long noncoding RNAs, often referred to as the Dark Matter of the genome, are gradually coming to light. Several recent papers dig deep to reveal surprisingly complex and diverse functions of these enigmatic molecules.

Noncoding RNAs (ncRNAs) differ from their better known counterpart messenger RNAs (mRNAs), by virtue of the fact that the sequence of bases contained within them do not encode proteins. They are generally divided into two classes based on an arbitrary length cutoff. Those under 200 nucleotides are usually referred to as short/small ncRNAs, including the microRNAs (miRNAs), and those greater than 200 bases are known as long noncoding RNAs (lncRNAs). Though several lncRNAs have been known for decades, the looming giant of lncRNAs was not fully exposed until genome-wide transcriptome studies revealed that approximately 10- to 20-fold more genomic sequence is transcribed to lncRNA than to protein-coding RNA. This potential treasure trove of thousands of lncRNAs has attracted intense scientific interest with the alluring possibility of finding new molecules and mechanisms that could shed light on organismal complexity. However, as lncRNA sequences are by definition noncoding, their potential functions are opaque to classical methods of making sense of genomic sequence. A rash of recent papers reveals that lncRNAs are important and powerful *cis*- and *trans*-regulators of gene activity that can function as scaffolds for chromatin-modifying complexes and nuclear bodies, as enhancers and as mediators of long-range chromatin interactions.

Cis-Acting lncRNAs; Local Gene Silencing

The most well-known lncRNA is *Xist,* which plays an essential role in X inactivation. During female development, *Xist* RNA is expressed from the inactive X and "coats" the X chromosome from which it is transcribed, leading to recruitment of Polycomb repressive complex 2 (PRC2), which trimethylates histone H3 at lysine 27 to silence transcription. Through its interaction with the X chromosome, *Xist* appears to create a nuclear compartment that excludes RNA polymerase II (RNAPII) (Chaumeil et al., 2006). Other lncRNAs such as *Air* and *Kcnq1ot1* also create repressive environments that may recruit and silence specific *cis*-linked gene loci by interacting with chromatin and targeting repressive histone modifiers (Nagano et al., 2008; Pandey et al., 2008) (Figure 1A). Though regulation of *Xist* transcription is not fully understood, it is clear that an overlapping antisense lncRNA, called *Tsix*, represses *Xist* expression in *cis*. Other lncRNAs such as *Xcite* and *RepA* also contribute to ensure that only one X chromosome is inactivated, by enhancing *Tsix* expression on the active X and upregulating *Xist* on the inactive X, respectively. Recent evidence suggests that both *Tsix* and *RepA* are able to bind PRC2 directly (Lee, 2010). Thus the major effector of X chromosome silencing, *Xist*, is itself controlled by a complex interplay of other *cis*-acting lncRNAs, some of which have been shown to function through recruitment of chromatin modification complexes.

Trans-lncRNA Activity; Hitting Multiple Targets

Unlike the *cis*-acting lncRNAs described above, a recent screen for lincRNAs (long intergenic noncoding RNAs) regulated by the tumor suppressor transcription factor p53 has revealed a lincRNA that targets silencing activity to multiple genes located throughout the genome (Huarte et al., 2010). In response to DNA damage, p53 triggers the activation or repression of numerous genes resulting in either cell-cycle arrest or apoptosis. Using inducible p53 cell systems, Huarte et al. showed that p53 regulates several lincRNAs, and one of them, *lincRNA-p21*, acts as a transcriptional repressor turning off multiple genes during the p53 response. Knockdown of either p53 or *lincRNA-p21* resulted in changes in expression of over 1000 genes, most of which were common to both knockdowns, and most of these resulted in gene derepression. The promoter of *lincRNA-p21* is directly activated by p53 binding in response to DNA damage. *lincRNA-p21* activity appears to trigger apoptosis rather than cell-cycle arrest. A search for factors that interact with *lincRNA-p21* identified heterogeneous nuclear ribonucleoprotein K (hnRNP-K), a component of a repressor complex that acts in the p53 pathway. hnRNP-K interacted with a 5′ domain of *lincRNA-p21* that was necessary but not sufficient to induce apoptosis, suggesting that other regions of the RNA are required to recruit other factors or target the complex to chromatin or both. Thus, *lincRNA-p21* is a *trans*-acting downstream repressor of multiple genes in the p53 pathway, potentially explaining how p53 can activate many genes while simultaneously repressing many others.

Scaffolds for Histone Modifiers

An important theme emerging from many of the latest studies is the ability of lncRNAs to bind chromatin modification complexes. Khalil et al. (2009) found that numerous lincRNAs are pulled down by RNA immunoprecipitation (RIP) of PRC2 and other

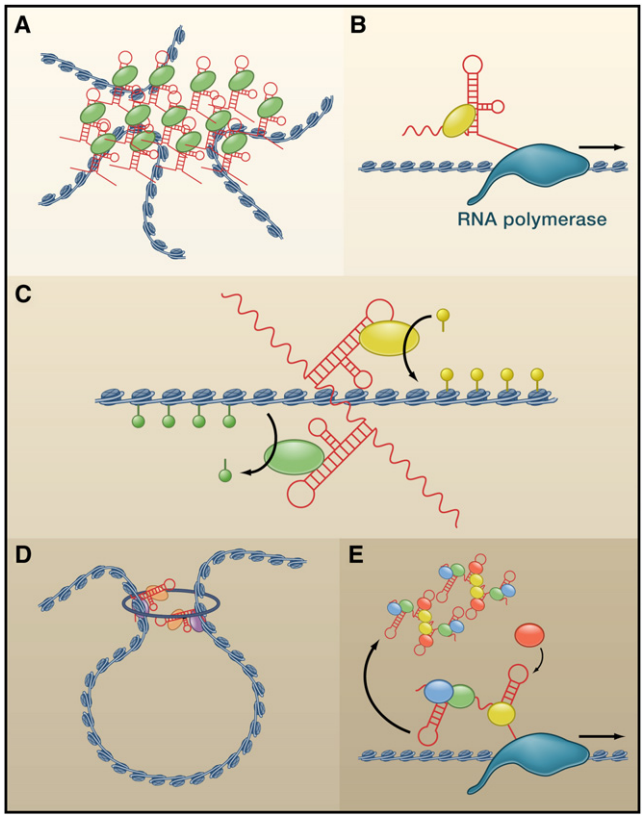

Figure 1. Modes of lncRNA Activity

(A) lncRNAs may nucleate chromatin from either the same or different chromosomes and create compartments enriched for chromatin modifier(s).

(B) lncRNA may cotranscriptionally recruit chromatin-modifying factors to specific chromosomal loci.

(C) One lncRNA may serve as a scaffold for multiple chromatin modifiers that alter different histone marks.

(D) Higher-order chromatin loops such as those mediated by CTCF and cohesin appear to involve lncRNA.

(E) lncRNAs generate the dynamic assembly of nuclear structures such as paraspeckles by serving as scaffolds for the associated proteins.

chromatin-modifying factors. Evidence of a functional union was bolstered by the finding that genes derepressed by siRNA knockdown of selected PRC2-associated lncRNAs were highly enriched in genes derepressed by disruption of PRC2 components (EZH2, SUZ12, and EED-1). Zhao et al. (2010) performed similar RIP experiments using PRC2 RIP-seq with nuclear RNA from mouse embryonic stem (ES) cells and identified thousands of PRC2-interacting RNAs, including large numbers of promoter transcripts, transcripts sense and antisense to known protein-coding genes, and a large proportion of imprinted gene loci. They provided in vitro evidence that the PRC2 complex may bind directly to RNA stem-loop structures via EZH2. Comparison of the PRC2 ''transcriptome'' to known PRC2-binding sites and bivalent domains (genomic regions with high H3K27me3 and H3K4me3) in ES cells revealed that many (~20%) bivalent domains contain at least one RNA, suggesting that RNAs may also recruit PRC2 to their sites of synthesis as well as to distal sites as described above (Figure 1B).

PRC2 is not the only histone-modifying complex found to bind to lncRNAs. The *HOTAIR* lncRNA is expressed from an intergenic region of the *HoxC* cluster and is necessary for PRC2 occupancy, H3K27me3, and silencing of the *HOXD* locus, located on a different chromosome (Rinn et al., 2007). Analysis of *HOTAIR* revealed that a 5′ end domain binds PRC2 and a 3′ end domain binds an LSD1 (H3K4me2 demethylase) containing complex (Tsai et al., 2010). Thus *HOTAIR* can act as a scaffold for these two distinct histone modification complexes and appears to target them to specific regions (Figure 1C) to remove the active histone modification H3K4me2, while methylating H3K27 toward a repressive mode.

What isn't clear from many of these studies is the precise mechanism by which these lncRNAs affect multiple genes. It is possible that they act as mobile scaffolds that target key complexes to multiple gene loci wherever they happen to be (Figure 1C). However, they may also function as organizing centers, performing the same functions by gathering multiple loci and factors into higher-order structures or discrete subnuclear locations or compartments (Figure 1A), such as described for polycomb bodies, or in a manner similar to that suggested for *Air* and *Kcnq1ot1*, which may be simplified modules for what happens repeatedly with *Xist* across the inactive X.

Organizing Enhancer Activity and Higher-Order Structures

Another potentially large lncRNA group is enhancer-related RNAs. Kim et al. (2010) found that many of the ~12,000 neuronal activity-regulated enhancers in the mouse genome are transcribed bidirectionally by RNAPII to yield noncoding enhancer RNAs (eRNAs). The expression level of eRNAs generally correlates with that of nearby protein-coding (target) genes, and in at least one example, eRNA expression required an intact target gene promoter, suggesting a reciprocal interaction between enhancers and promoters during promoter activation. De Santa et al. (2010) also investigated transcription of enhancers. They focused on RNAPII-binding peaks and noncoding transcription outside of protein-coding genes during macrophage activation and matched these extragenic sites with distinct chromatin signatures characteristic of enhancers. They found large numbers of RNAPII-bound enhancers and eRNAs, suggesting that transcription of enhancers may be a general feature. However, the possibility that eRNAs are biproducts of target gene activation could not be excluded, as it was not confirmed that they play an essential role.

Evidence that lncRNAs themselves may have enhancer function was put forward by Ørom et al. (2010). They used siRNA knockdown to test the possible function of several lncRNAs, all of which were located further than 1 kb from known protein-coding genes. Importantly, these lncRNA loci bore the chromatin signatures of transcribed protein-coding gene loci (H3K4me3 at the 5′ end and histone H3 lysine 36 trimethylation downstream), suggesting that they are not enhancer elements, which are characterized by H3K4 monomethylation. Knockdown of these lncRNAs resulted in corresponding decreases in expression of neighboring protein-coding genes. They designated seven activating ncRNAs, *ncRNA-a1* through *ncRNA-a7*, that appear to enhance the expression of neighboring protein-coding genes.

Whether these ncRNAs work with other factors is not known, but it's tempting to speculate that they function in a manner akin to the above-mentioned silencing lncRNAs, as coactivators that recruit positive-acting factors. Conversely, they may work by physically juxtaposing a putative partner factor with the promoter region of the target gene by long-range loop formation.

A couple of recent papers further suggest mechanisms along these lines. Yao et al. (2010) showed that the DEAD-box RNA helicase p68 (DDX5) and its associated lncRNA, *SRA* (steroid receptor RNA activator), form a complex with CTCF. CTCF binds to specific genomic binding sequences and plays an important role in transcriptional insulation and long-range physical interaction with other CTCF sites. These interactions are mediated by the ring-like cohesin complex that appears to use chromatin-bound CTCF as a binding platform (Figure 1D). CTCF's insulator function is dependent on p68 and *SRA*, as depletion of either mitigates CTCF-mediated insulation between *IGF2* and its long-range enhancer at the *IGF2/H19* locus. p68 binds both *SRA* and CTCF, and *SRA* stabilizes binding between CTCF and cohesin. Depletion of either p68 or *SRA* did not affect CTCF binding to its genomic sites but reduced the presence of cohesin at these sites.

Another example involves the homeodomain transcription factor genes *Dlx-5* and *Dlx-6* and an intergenic ultraconserved region. Ultraconserved regions are noncoding genomic sequences of over 200 bases that are 95%–100% conserved among several species, from the fish to human. The startling degree of conservation of these noncoding sequences among such distant species has sparked the suggestion that they constitute fundamental vertebrate regulatory elements. The *Dlx-5/6* ultraconserved region is transcribed as part of the *Evf-2* lncRNA in response to sonic hedgehog signaling in the developing telencephalon. *Evf-2* has transcriptional regulatory activity mediated through the ultraconserved sequences at its 5′ end, which forms a complex with the Dlx-2 transcription factor (Feng et al., 2006). The *Evf-2*/Dlx-2 complex has been proposed to affect transcriptional activity, possibly by stabilizing the association with the *Dlx-5/6* enhancer to activate *Dlx-5/6* gene expression. Assuming that the enhancer then works via looping to the distal promoters, the net results may be stabilized factor binding at the enhancer-promoter complex and potentially stability of the higher-order complex. Bond et al. (2009) have presented evidence that *Evf-2* also recruits MECP2 to DNA and that this balancing of a positive and negative factor regulates *Dlx-5/6* enhancer activation of *Dlx-5/6* gene expression.

The most obvious connection between these positively acting lncRNAs and some of the above-mentioned silencing lncRNAs is the fact that they appear to function locally to affect *cis*-linked gene loci. However, further examination of the *Xist* regulation paradigm has revealed a new, potentially *trans*-acting activator lncRNA. The *Jpx* lncRNA is located upstream of the *Xist* transcription unit and positively regulates *Xist* expression (Tian et al., 2010). Deletion or knockdown of *Jpx* led to failure of *Xist* upregulation and *Xist* coating of the X chromosome during differentiation of female ES cells, whereas it had no effect in male cells. Surprisingly, deletion of a single copy of *Jpx* in female ES cells did not result in preferential inactivation of the wild-type chromosome. Such skewing of the normally random X inactivation process usually occurs when *Xist* expression is disrupted on one of the X chromosomes. Instead, *Jpx* deletion heterozygotes had less than the expected 50% of residual *Jpx* RNA and showed a dramatic failure in *Xist* coating and X inactivation. *Xist* expression and X inactivation could be rescued by a *Jpx* transgene located on another chromosome, indicating that *Jpx* can exert its effects in *trans*. Exactly how *Jpx* augments *Xist* expression or indeed how the two *Jpx* alleles cooperate to control their expression in female cells are not known. The fact that *Jpx* is also upregulated during male ES cell differentiation without consequent upregulation of *Xist* suggests that it does not work alone. Chureau and colleagues (2010) report that *Ftx*, another conserved lncRNA located just downstream of *Jpx*, also positively affects *Xist* expression. Like *Jpx*, *Ftx* partially escapes X inactivation, meaning that it is transcribed from both the active and inactive X chromosomes. However, unlike *Jpx*, *Ftx* is upregulated specifically in female cells at the time of *Xist* upregulation and X inactivation. Whether *Ftx* can also function in *trans* is not known; however the picture is further complicated by the fact that *Ftx* hosts several miRNAs within its introns, one of which (miR-421) potentially targets ATM. ATM plays a central role in genome integrity by promoting double-strand break repair, and disruption of its function leads to silencing defects on the inactive X chromosome (Ouyang et al., 2005). Importantly, neither *Jpx* nor *Ftx* appear to function merely as negative regulators of *Tsix*. Together with *Tsix*, *RepA*, and *Xcite*, they begin to flesh out a complex and elaborate regulatory network of multiple lncRNAs that affect *Xist* expression and X inactivation though *cis* and *trans* silencing and activation mechanisms.

lncRNA in Genomic Reprogramming

With all the varied and powerful functions of lncRNAs, it is perhaps not surprising that they have been implicated in global remodeling of the epigenome and gene expression during reprogramming of somatic cells to induced pluripotent stem cells (iPSCs). Loewer and colleagues (2010) looked for lincRNAs that are specifically upregulated in human iPSCs compared to the cell of origin and identified a subset of those that are elevated in iPSCs compared to ES cells, reasoning that their increased expression may promote reprogramming. They found that iPSC-enriched lincRNA loci are bound by the key pluripotency transcription factors OCT4, SOX2, and NANOG, and knockdown of OCT4 led to downregulation of the lincRNAs, suggesting that their expression is directly regulated by the pluripotency factors. They focused on two of these lincRNAs, *lincRNA-RoR* and *lincRNA-SFMBT2*, which showed the strongest response to OCT4 knockdown, and investigated their potential role in reprogramming by knocking them down in fibroblasts and assessing iPSC colony formation induced by infection with viruses expressing the pluripotency factors. Knockdown of *lincRNA-RoR* resulted in a significant decrease in iPSC colony formation compared to control cells, indicating that it plays a role in iPSC derivation. This idea was further supported by the finding that cells stably overexpressing *lincRNA-RoR* were 2-fold more efficient in iPSC colony formation. To gain insight into pathways affected by *lincRNA-RoR*, they assessed gene expression by microarray and found that knockdown of *lincRNA-RoR* led to upregulation of genes involved in the p53 response, the response

to oxidative stress and DNA-damage-inducing agents, and cell death pathways, suggesting that *lincRNA-RoR* plays a role in promoting iPSC survival.

lncRNAs as Nucleators of Nuclear Structures

In a slightly different twist on the emerging theme of lncRNAs acting as scaffolds for factors that target chromatin and gene expression, recent live-cell results show that lncRNAs can also act as platforms for the assembly of dynamic nuclear structures (Figure 1E).

Paraspeckles are discrete ribonucleoprotein bodies found in mammalian cell nuclei, implicated in nuclear retention of hyper-edited mRNAs. Mao et al. expressed fluorescently tagged paraspeckle-associated fusion proteins in cells with an inducible *Men* ε/β lncRNA, the RNA component of paraspeckles (Mao et al., 2011). The *Men* ε/β lncRNAs themselves were tagged with an array of hairpin-binding sites for the MS2 viral coat protein, which was fused to EYFP to allow visualization of the nascent *Men* ε/β transcripts. They showed that paraspeckle-associated proteins were rapidly recruited and assembled on the *Men* ε/β lncRNAs as they were being transcribed and that these assembled structures persisted near the nuclear site of transcription, as has been shown for endogenous *Men* ε/β-containing paraspeckles. Also, like endogenous paraspeckles, the induced structures effectively retained specific mRNAs, suggesting that they were functional. The authors showed that maintenance of paraspeckle structures was dependent on active transcription of the *Men* ε/β lncRNAs.

Temporary and reversible blocking of transcription led to disassembly of paraspeckle components, whereas reversal of the transcriptional block resulted in reassembly of paraspeckle proteins on nascent *Men* ε/β lncRNAs only, not on mature *Men* ε/β. Shevtsov and Dundr (2011) went a step further and showed that several types of nascent RNAs (noncoding and protein-coding) can trigger assembly of various nuclear bodies by serving as scaffolds for accumulation of specific proteins, accentuating the capability of RNAs to act as modular scaffolds for the rapid assembly of multiple components.

Making Sense of the Nonsense

These exciting new functions and potential mechanisms of lncRNAs, combined with the unexplored enormity of noncoding transcripts in higher organisms, suggest that many new roles in gene control and genome and nuclear organization are likely to be uncovered. How many of the remaining thousands of lncRNAs will be functional is difficult to say, but it is now clear that it is not all junk, derived from promiscuous transcription. A strong emerging theme is the apparent ability to function as scaffolds for regulatory factors that then target those factors to gene loci, which might be accomplished in several ways. Some lncRNAs may recruit chromatin-modifying complexes to the site of their transcription, whereas others target chromatin modifiers to distant loci. Formation of a nuclear compartment enriched with chromatin modifiers or other regulatory factors may enable efficient control of multiple loci simultaneously; however, it is also possible that lncRNAs act as mobile scaffolds that target individual genes in a manner analogous to a transcription factor. In addition, lncRNAs are involved in forming higher-order chromatin loops and can act as scaffolds for the assembly of proteins involved in formation of nuclear structures and functional nuclear subcompartments. It appears that dynamic protein assembly onto nascent lncRNA seeds is a common theme, suggesting that synthesis of new lncRNAs could rapidly form regulatory complexes with the potential to target ubiquitous regulatory factors to implement diverse gene expression patterns during differentiation, development, and reprogramming.

REFERENCES

Bond, A.M., Vangompel, M.J., Sametsky, E.A., Clark, M.F., Savage, J.C., Disterhoft, J.F., and Kohtz, J.D. (2009). Nat. Neurosci. *12*, 1020–1027.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). Genes Dev. *20*, 2223–2237.

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2010). Hum. Mol. Genet. *20*, 705–718.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). PLoS Biol. *8*, e1000384.

Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. (2006). Genes Dev. *20*, 1470–1484.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). Cell *142*, 409–419.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Proc. Natl. Acad. Sci. USA *106*, 11667–11672.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Nature *465*, 182–187.

Lee, J.T. (2010). Cold Spring Harb. Perspect. Biol. *2*, a003749.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Nat. Genet. *42*, 1113–1117.

Mao, Y., Sunwoo, H., Zhang, B., and Spector, D. (2011). Nat. Cell Biol. *13*, 95–101.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). Science *322*, 1717–1720.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Cell *143*, 46–58.

Ouyang, Y., Salstrom, J., Diaz-Perez, S., Nahas, S., Matsuno, Y., Dawson, D., Teitell, M.A., Horvath, S., Riggs, A.D., Gatti, R.A., et al. (2005). Biochem. Biophys. Res. Commun. *337*, 875–880.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Mol. Cell *32*, 232–246.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Cell *129*, 1311–1323.

Shevtsov, S.P., and Dundr, M. (2011). Nat. Cell Biol. *13*, 167–173.

Tian, D., Sun, S., and Lee, J.T. (2010). Cell *143*, 390–403.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Science *329*, 689–693.

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., and Felsenfeld, G. (2010). Genes Dev. *24*, 2543–2555.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Mol. Cell *40*, 939–953.